R E P O R T   R E S U M E S

ED 012 003

AL 000 031

PROGRESS REPORT ON MANDARIN TONE STUDY.
BY- BROTZMAN, ROBERT
OHIO STATE UNIV., COLUMBUS, RESEARCH FOUNDATION
REPORT NUMBER POLA-8-1 PUB DATE APR 64
EDRS PRICE MF-$0.09 HC-$1.52 38P.

DESCRIPTORS- *MANDARIN CHINESE, *SYNTAX, *DESCRIPTIVE LINGUISTICS, *STRUCTURAL ANALYSIS, *ACOUSTIC PHONETICS, COMPUTATIONAL LINGUISTICS, PROJECT ON LINGUISTIC ANALYSIS, TONE SANDHI, PITCH CONTOUR, TUPLES, COLUMBUS

THE FORM OF A GENERATIVE SYSTEM FOR FUNDAMENTAL VOICE FREQUENCY IN MANDARIN CHINESE IS DISCUSSED, AND SEVERAL ASSUMPTIONS ARE MADE IN ORDER TO REDUCE THE PROBLEM TO A MANAGEABLE SIZE. DATA WAS GATHERED FROM TWO SPEAKERS BY HAVING THEM READ A PREPARED LIST WHICH CONTAINED ISOLATED WORDS, TWO-TUPLES, AND THREE-TUPLES IN ALL POSSIBLE COMBINATIONS OF THE FOUR BASIC TONES. THE DATA GATHERING SYSTEM UTILIZED A MODIFIED "VOCODER" PITCH EXTRACTOR, A TWO-CHANNEL GRAPHICAL RECORDER, AND AN ELECTRONIC DIGITAL COMPUTER WHICH WAS USED TO PLOT OUT THE PITCH CURVE AND ALSO TO EXTRACT VARIOUS PARAMETERS OF THE PITCH CONTOUR. EXAMPLES ARE GIVEN OF THE FOUR BASIC TONES OF MANDARIN, AND, ON THE BASIS OF THE DATA GATHERED, A SET OF RULES IS PROPOSED WHICH WOULD ACCOUNT FOR THE TONE SANDHI AND CERTAIN ALLOTONIC CHANGES EVIDENT IN MANDARIN SPEECH. (IT)

RF Project 1685
Report 2

THE OHIO STATE UNIVERSITY

RESEARCH FOUNDATION

1314 KINNEAR ROAD COLUMBUS, OHIO 43212

Project on Linguistic Analysis
Report No. 8

PROGRESS REPORT ON MANDARIN TONE STUDY

Robert Brotzman
Division of Linguistics

[redacted]

[redacted]

[redacted]

NATIONAL SCIENCE FOUNDATION
Grant No. GN-174

Progress Report on Mandarin Tone Study

Robert Brotzman

I. Introduction

A grammatical analysis of a language may be thought of as divided into two parts, the syntax and the phonology. The lexicon might be assumed to be a separate entity or a part of the syntax. In the phonology of a language such as Mandarin, where the pitch pattern carries lexical information, the phonology must have one section which contains the rules for generating this pattern from information provided by the syntax and the lexicon. The phonological rules have been classified into six major types as follows: [1]

(1) RS rules--those which add the retroflex suffix and effect a change on the final of the syllable;
(2) SV rules--those which change vowels in certain positions into glides;
(3) RD rules--those which reduplicate syllables;
(4) SA rules--those which assign stress;
(5) TS rules--those which provide for the correct tone sandhi;
(6) TP rules--those which place the tone on the proper portion of the syllable.

This study is an attempt to discover something about the nature of the TS rules, and hopefully to formulate some of them. Assuming for the moment that the lexicon is a part of the syntax, the block diagram shown in Figure 1 might be assumed to operate in the generation of a pitch pattern for Mandarin.

For our purposes here, we will accept the traditional analysis of Mandarin having four distinctive tones, i.e. 1 - high level, 2 - high rising, 3 - low dipping, and 4 - high falling. "Sandhi" rules are those which replace one tone by another, whereas the "allotonic" rules are those which specify the environmental influence on the actual pitch curve of a tone. Although the stress rules are located after the sandhi and allotonic rules, it is assumed that the stress positions have already been specified, so that the stress rules consist only of the physical specifications of the pitch curve which are necessary to produce a sentence which <u>sounds</u> stressed in the places previously specified.

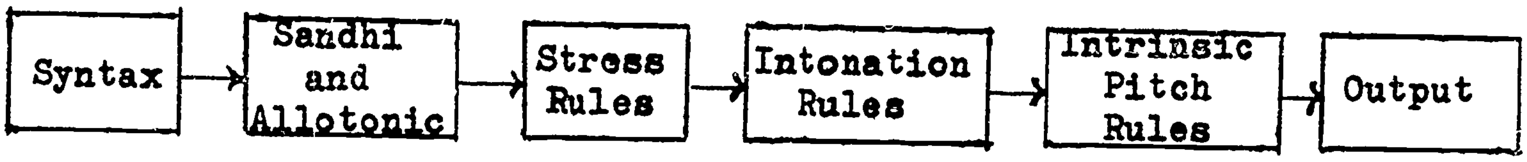


Figure 1. Block Diagram of Pitch Curve Generator

There are reasons for believing that some types of stress must be specified prior to the generation of the sentence structure, or at least during this generative procedure. As an example of the necessity of knowing whether or not any part of a sentence is stressed, consider the following examples, one in Mandarin and one in English:

'I want to go too.' 'Wǒ yě yào qù.' (I also/too want to go.)

In both of these cases the structure is permissible if stress is present, but is not permissible when none of the words is stressed. A further reason for desiring stress positions to be marked before the phonology operates will be illustrated in Section IV, where the operation of a sandhi rule is governed by the presence or absence of stress. Stress will therefore be assumed to be marked within the syntax component. The output of the syntax is assumed to be strings of morphemes with their structural descriptions, plus certain phonological information.

During the initial stages of this investigation it will be assumed that the pitch pattern alone is the parameter which carries the information types indicated by Figure 1. Although this is certainly not true, it provides a simplification which brings the problem down to a manageable size to begin with. Later in this study, tests will be made to determine the influence of other factors, primarily intensity, used by listeners to determine the tone of a word.

As can be readily observed in the block diagram of Figure 1, the pitch pattern is carrying information of several types. Omitting the effect of intrinsic pitch,[2] since it carries no information, we have the single parameter of fundamental frequency (by assumption) carrying the following three loads:

(1) lexical information;

(2) stress information;

(3) intonation pattern.

Since this information is carried in a single parameter, we must expect to have interaction among the different types of information carried, to produce a sequence which has a unique or nearly unique interpretation to a native speaker.

As yet, even the characteristic pitch curves of Mandarin words spoken in isolation are not all known, or at least not all agreed upon. A good deal of the difference of opinion is due, no doubt, to the fact that nearly all the work along this line in the past was done, of necessity, in an impressionistic manner. [3] The tool that is probably most commonly used now in studying the fundamental frequency of voice is the Sona-Graph. Narrow band spectrograms give the researcher accurate, continuous pitch contours from which to extract data.

In this study the Sona-Graph was used very little, for two reasons. First, the collection of data by this method takes an exceedingly long time. Second, the Sona-Graph does not lend itself readily to an automatic data processing system which might be used in automatic recognition of the spoken tone. The latter reason has considerable bearing on the decisions not to use a Sona-Graph, since it is intended that recognition tests using the generative rules in reverse be used to aid in the understanding of how these generative rules should be written. While the reverse of the generating procedure can hardly undo conversions of one tone into another due to tone sandhi operating, since there is a many-to-one relationship between input and output, it is felt that development of recognition procedures will provide a valuable first step to understanding the generative system. Since the recognition procedures will incorporate information gathered from real speech, even a highly accurate recognition procedure can only take into account what has already been observed. Therefore, upon formulation of an acceptable recognition scheme, it will still be necessary to synthesize tone patterns and have them judged by native speakers, in order to set reasonably accurate limits to the permissible variations of pattern.

II. Procedure and Equipment for Data Gathering

A set of sixteen Mandarin words was selected, which had the following properties:

(1) All words had the same vowel, /a/.

(2) The sixteen words could be broken down into four groups, each group representing a single consonant-vowel sequence, and differing only in that the four basic tones of the language were present in each group.

As was mentioned by Chao, [4] the gathering of data on a pitch curve is complicated by the tendency of the informant to impose sentence intonation on polysyllabic groups, and to shift key when reading sets of isolated words. In this experiment the isolated words were merely spaced in time. It might be better in the future to use a carrier sentence for the target words, in spite of the fact that the tones preceding and following the target word will affect the pitch contour. The choice is between accepting the variations present in isolated words and accepting a uniform deformation due to the tone's environment.

The original list of sixteen words was read twice by each of the two informants, [5] with a pause of several seconds between each word. The informants then read two-tuple combinations of each set of segmentally identical words in all possible combinations, and finally they read three-tuple combinations of the segmentally identical sets in all possible combinations. In addition to this, each informant read a set of expressions which included "neutral tones," and a set of sentences in which the sentence was first read without stress and then reread, stressing a different word each time. All written lists were prepared in characters.

The readings were made in a sound-proof recording booth, and recorded on an Ampex 354 tape recorder at 7-1/2 ips, using "Scotch 175" recording tape and an Altec 683A dynamic microphone. The recordings were then replayed on the Ampex recorder, and the electrical signal was fed into a modified "Vocoder" pitch extractor circuit. [6] The pitch signal was then sent to one channel of a two-channel graphical recorder, the

other channel being fed the overall amplitude curve. The graphical recorder was set at 125 mm/sec, and the frequency scale was calibrated at 5-minute intervals, using a signal generator and a frequency counter. These graphical tapes were then edited by hand to mark word boundaries, and to write in word and tone. For the words with unvoiced initials, the determination of word boundaries was no problem, since the pitch contour started at the onset of voicing. For the words with a voiced initial the boundary was set by observing both the pitch contour and the synchronous amplitude curve, since a voiced consonant has a lower intensity than the vowel /a/.

Pitch and amplitude values were then read from the tape at 40-msec intervals during each word, and recorded on punched cards along with the information on tone and environment. These cards were then processed by an IBM 7090 computer, to convert the scale on the graphic tapes into a frequency scale for pitch, and to plot: pitch on a frequency versus time scale; amplitude on a relative scale versus time; and normalized frequency on a ratio versus time scale. The normalized frequency was obtained by dividing the frequency at each point by the value of the frequency at the initial point of the most recent Tone 1. After this output was studied to determine what parameters might be useful in specifying tones, the cards were again run through a computer, to extract the following information:[7]

(1) The lexical tone assigned by the vowel, as well as that of the preceding and/or following words.
(2) Average frequency of the vowel.
(3) Frequency of a local minimum if one existed.
(4) Slope toward the local minimum from beginning of the vowel.
(5) Slope away from the local minimum to end of the vowel.
(6) Initial and final frequencies, and the difference between them.
(7) Concavity, whether up or down, and the degree of concavity.
(8) Location of the minimum in number of 40-msec units from beginning of the vowel.
(9) Number of 40-msec units in the vowel duration.

(10) Average slope of the pitch curve from beginning to end of the vowel.

(11) The starting slope of the curve, averaged over the first two 40-msec intervals.

These data were then manipulated by hand to determine the effects that tone and tone environment had on the various parameters.

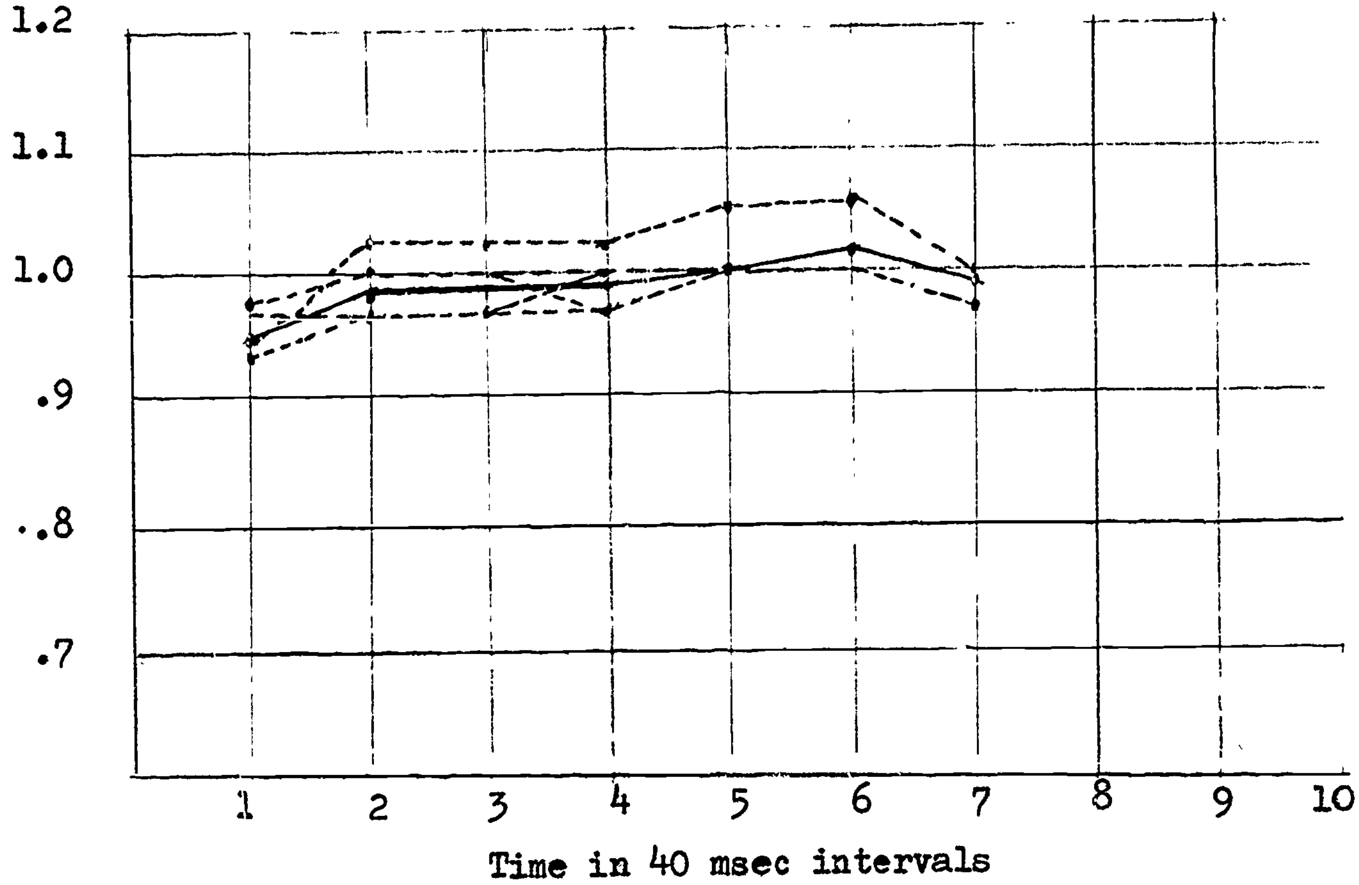
III. Isolated Words

In Figure 2 are shown some patterns of each of the four basic tones of Mandarin. Throughout this report, patterns represented by dotted points will indicate the female informant; patterns represented by dashed points will indicate the male informant. Both curves are normalized with respect to the average frequency of the informants' Tone 1. The solid lines represent the average of the curves on each graph, and approximate the characteristic pattern of the tone. These curves were selected at random from the isolated words spoken that did not have voiced initial consonants. The words with voiced initial consonants were eliminated from this graph because of the slightly different pitch pattern which is produced. Figure 3 shows the pitch patterns of words having voiced initials, with these voiced initials included in the pitch patterns. As can be noted in Figure 3, the voicing during the consonant is approximately level, and tends toward a middle position in the pitch range. It was assumed that the boundaries between consonants and vowels would be known by other methods, and in this study no attempt has been made to specify the pitch contour during the period of voicing for consonants. Naturally this must be done before one can completely specify a pitch curve.

As can be readily seen in Figure 2, the pitch patterns for the four basic tones spoken in isolation are quite distinct. Differentiations can be made on the basis of the following observations:

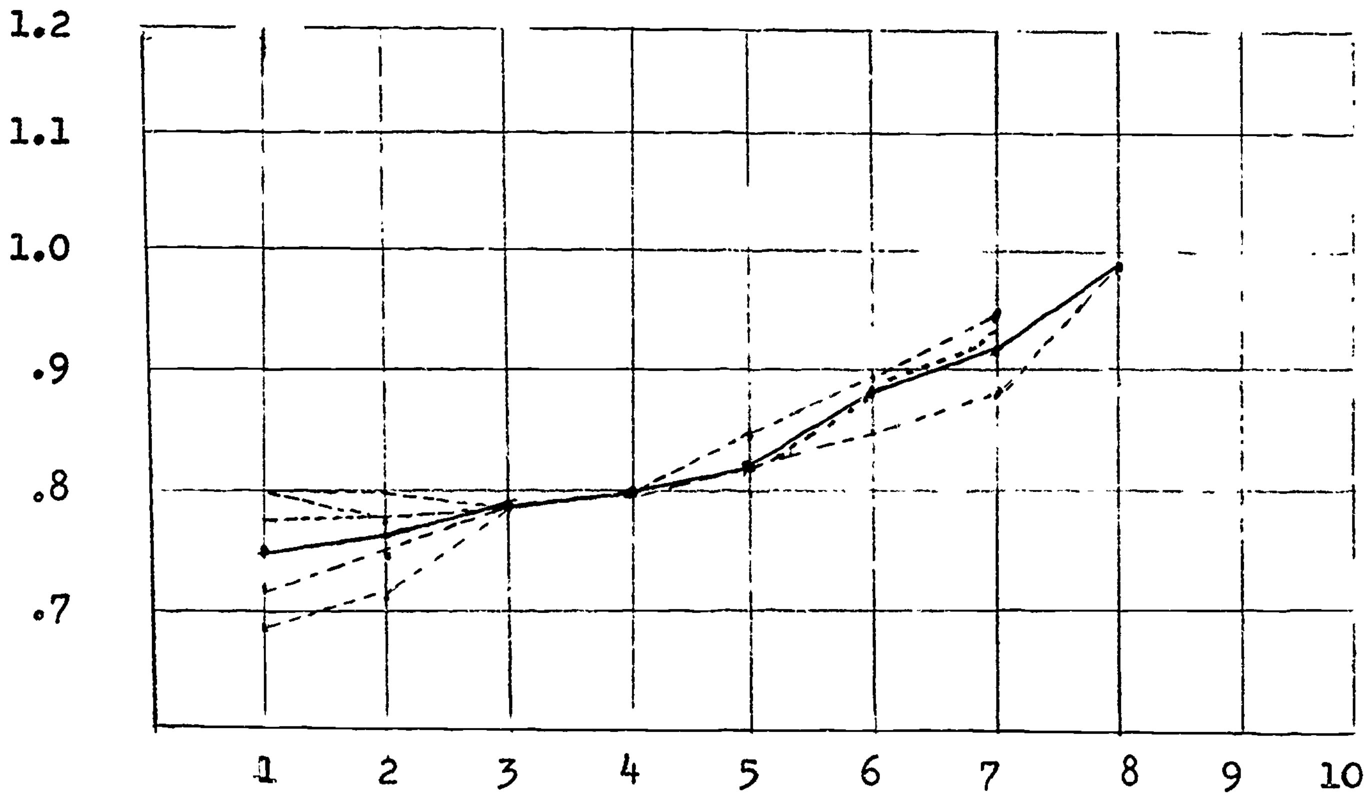
(1) The average frequency of Tone 1 did not intersect with the average frequencies of any of the other tones. It is the highest of the averages.

(2) The average frequency of Tone 3 did not intersect with the average frequencies of any of the other tones. It is the lowest of the averages.



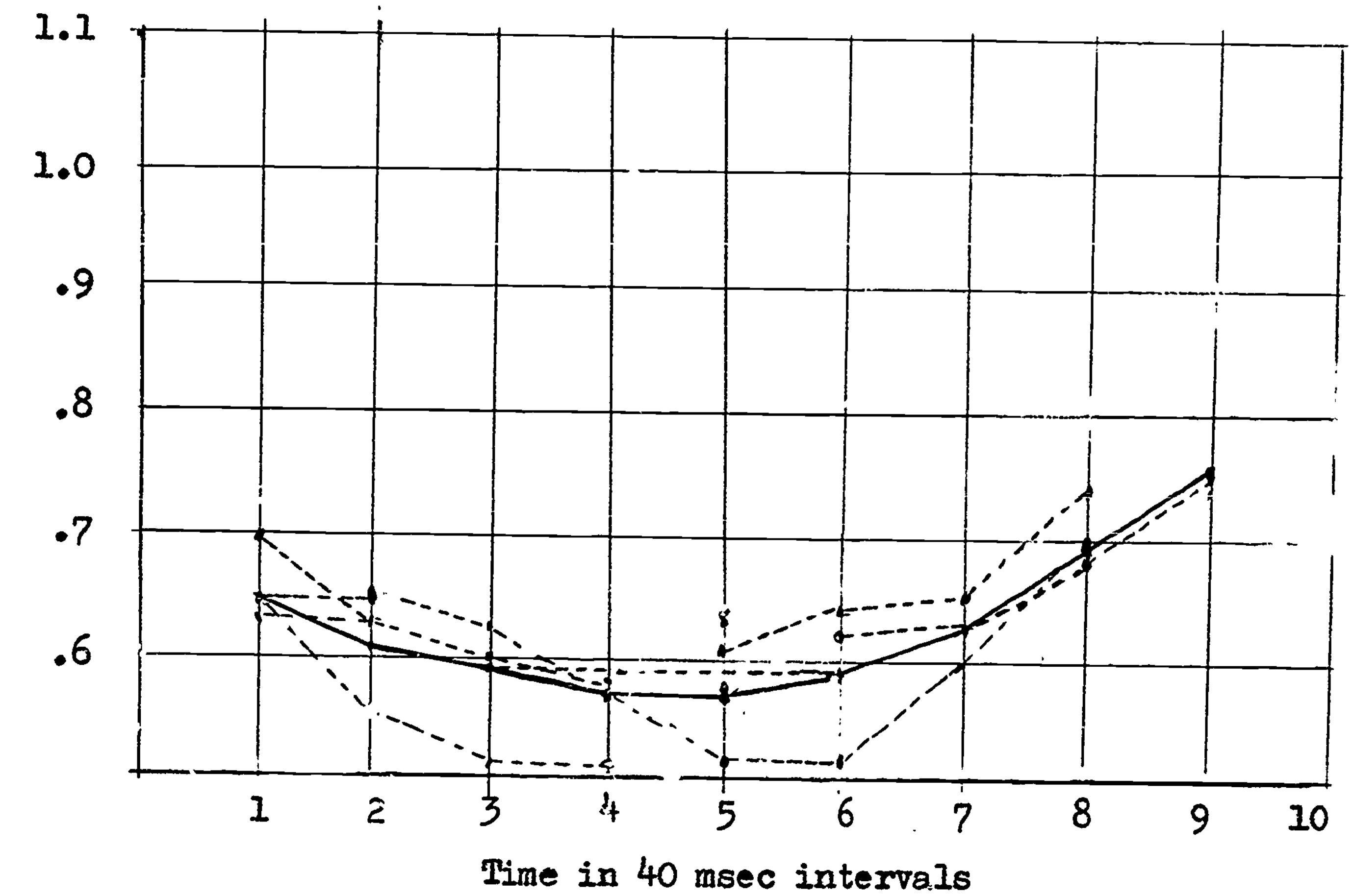
Tone #1

3 male and 3 female in each tone
male = -----
female = ...
average = ———

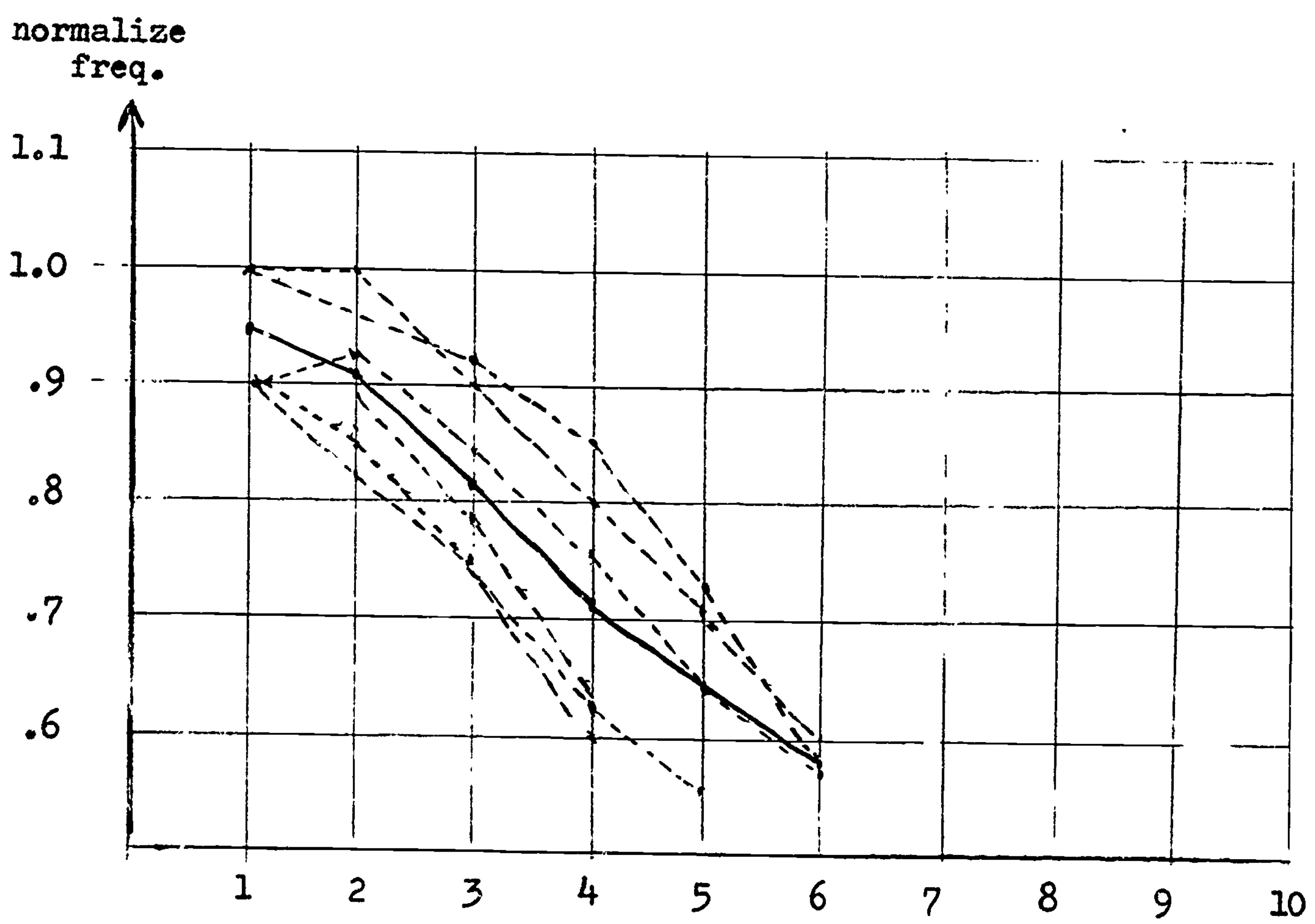


Tone #2

Figure 2 (Part A) Words with Unvoiced Initial Consonants

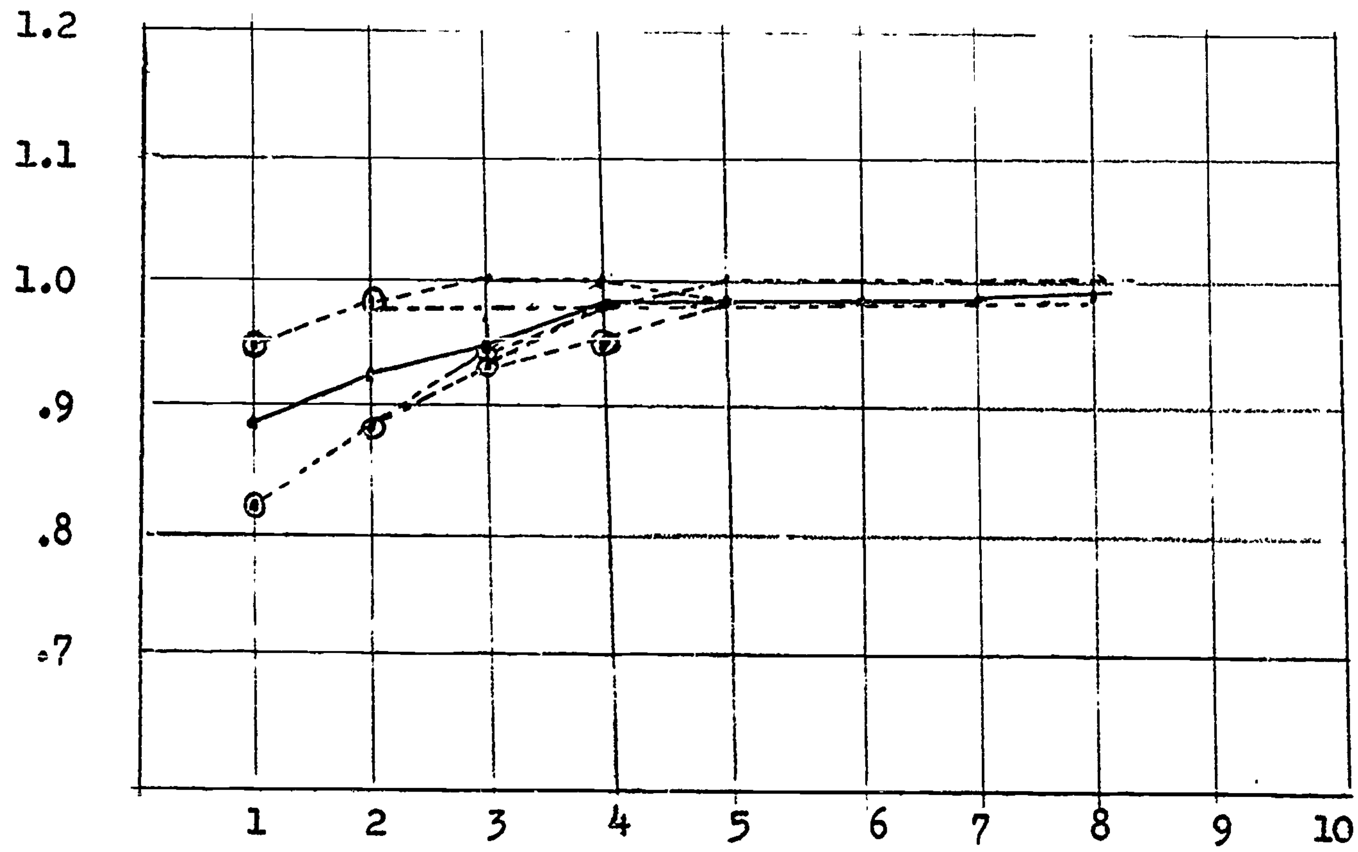


Tone #3



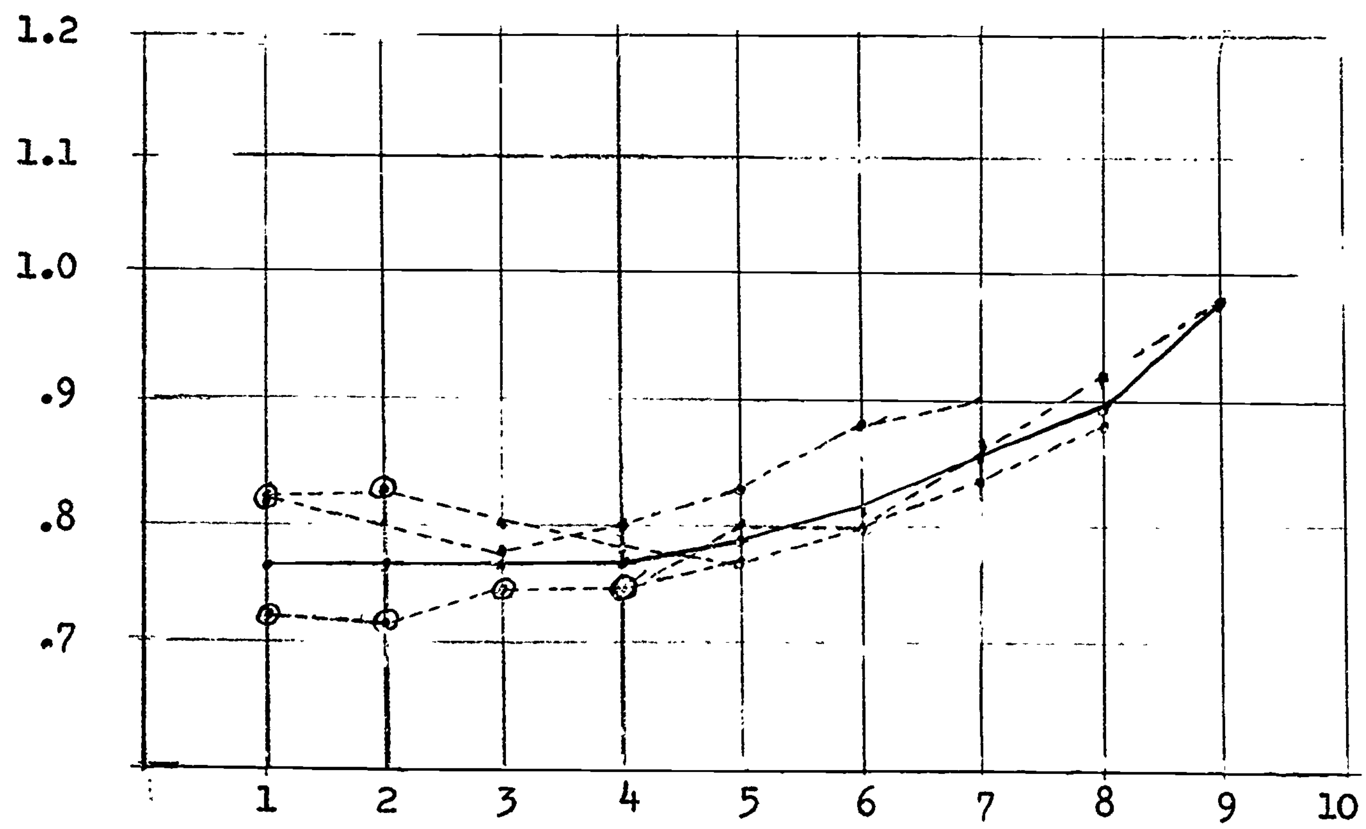
Tone #4

Figure 2 (Part B) Words with Unvoiced Initial Consonants



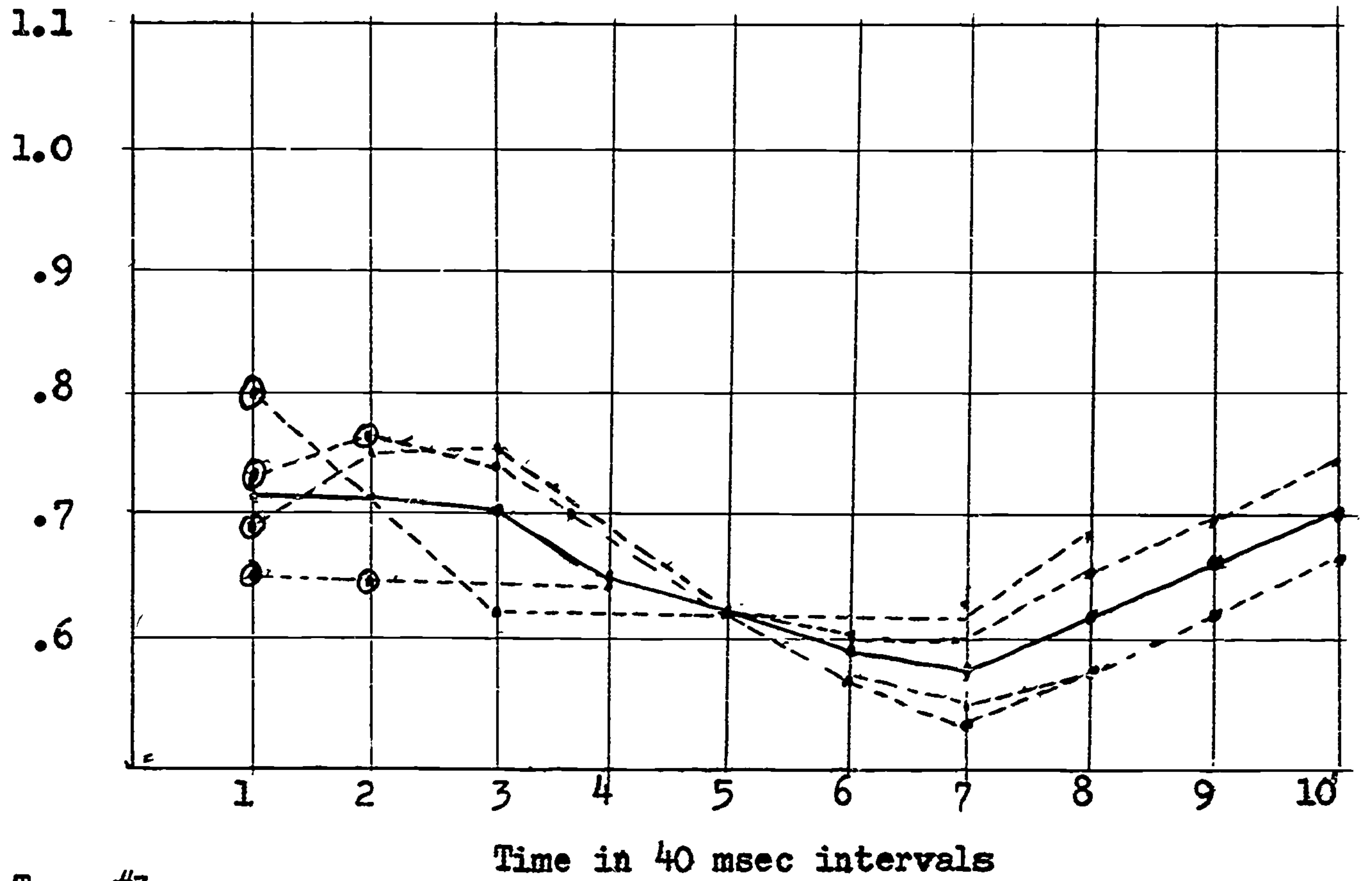
Tone #1

Time in 40 msec intervals
(Circled points indicate voicing in the consonant)

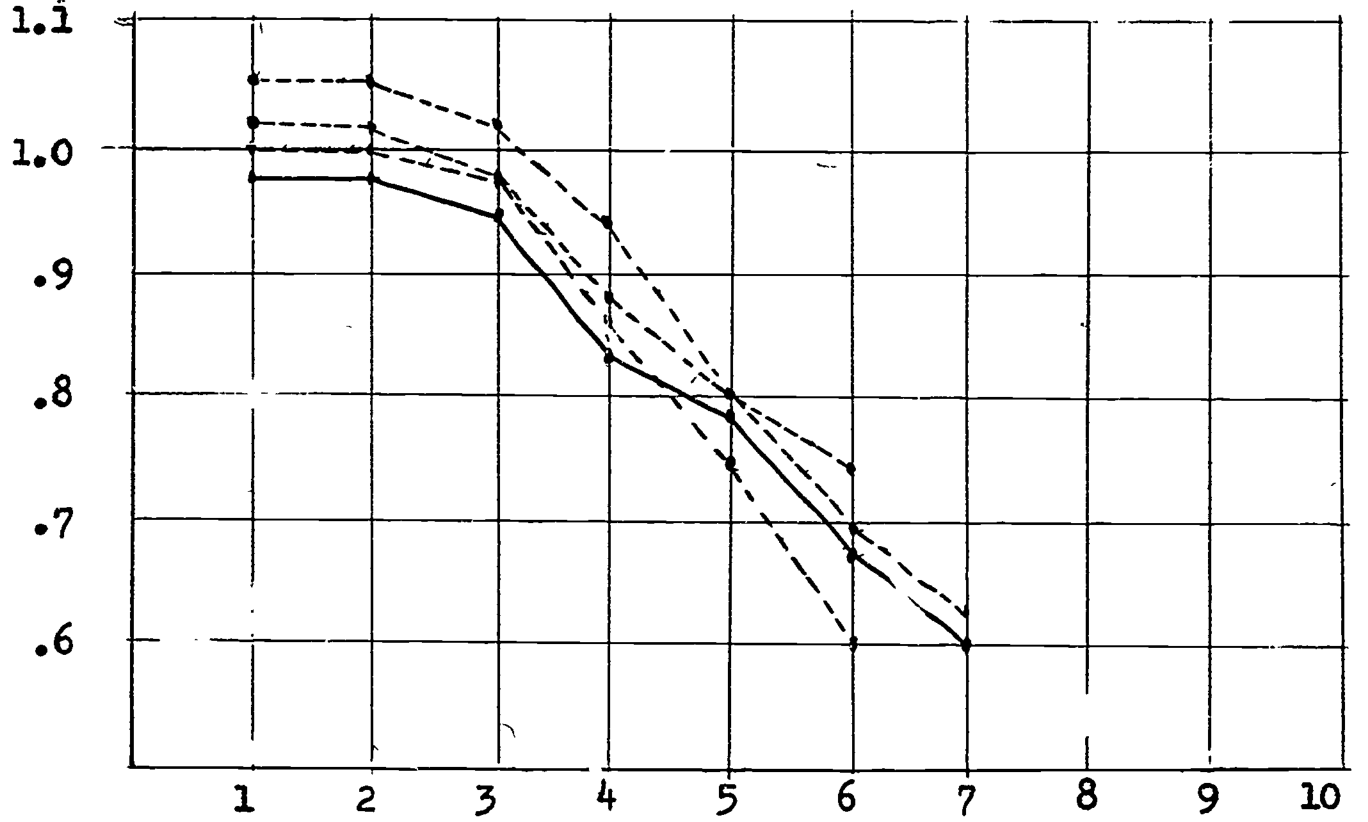


Tone #2

Figure 3 (Part A) Words with Voiced Initial Consonants (ma)



Tone #3



Tone #4

Figure 3 (Part B) Words with Voiced Initial Consonants (ma)

(3) Tones 2 and 4, although occupying the same area of average frequency, i.e. the middle area below Tone 1 and above Tone 3, are easily distinguished on the basis of average slope. The slope of Tone 2 is always positive, and that of Tone 4 is always negative. Due to the small sample from which the above observations were made, we can hardly assume that the average frequencies of Tones 1 and 3 cannot intersect with those of Tones 2 and 4. It therefore seems advisable to set up an intermediate area at the upper and lower ends of the Tone 2, Tone 4 average frequency area, and made decisions on the basis of other parameters in these ranges. Figure 4 presents a decision procedure which is designed to resolve ambiguities which might be encountered. At present this procedure had not been tested by a computer, but it will be in the near future. Until such time as the recognition procedure shall be developed to an acceptable state of accuracy, no attempt will be made to formulate rules for the physical pitch curves of the isolated words. Appendix 4 gives average, maximum, and minimum values for the pertinent parameters of each tone spoken in isolation.

IV. Words in Sequence

As soon as we start dealing with words in sequence we face the problems of tone sandhi and allotonic change. It must be understood that the sandhi and allotonic changes discussed below do not apply to the reduplicated expressions in Mandarin. These expressions have an unusual behavior and must be handled separately, probably in the syntax. There is also a small set of words which have unusual tone behavior. These words are Bù, Yī, Chī, and Bā (不, 一, 七, 八). The tones marked are the normal citation tones. The most widely known sandhi change in Mandarin is the changing of a Tone 3 into a Tone 2 when it precedes a Tone 3, i.e. 33 ---→ 23. Even this change was open to debate, since some writers [8] claimed that the Tone 3 did not really become a Tone 2 but rather a new tone, called Tone 5. While there is some evidence that the Tone 3 does not become a perfect Tone 2 in this environment. Tone 3 is perceived as, and indistinguishable from, a Tone 2 by native speakers, as shown by the experiment of K-P. Li. [9] The

only other tone sandhi commonly mentioned is that of Tone 2 becoming a Tone 1 in certain environments. As can be deduced from the examples presented by Chao in his Mandarin Primer,[10] Tone 2 becomes a Tone 1 when it follows either Tone 1 or Tone 2 and does not precede a pause. In addition to some rules for neutral tones, with which we are not dealing at the present time, Chao also mentions two allotonic changes; that of a Tone 4 not falling as low as usual when it precedes another Tone 4, and that of a Tone 3 not rising when it precedes another tone. Hence we could write Chao's sandhi and allotonic changes as follows, with all rules repeatable. The mark ( ) indicates that any of the tones inside may appear in this position.

(1) $33 \longrightarrow 23$

(2) $\begin{pmatrix}1\\2\end{pmatrix} 2 \begin{pmatrix}1\\2\\3\\4\end{pmatrix} \longrightarrow \begin{pmatrix}1\\2\end{pmatrix} 1 \begin{pmatrix}1\\2\\3\\4\end{pmatrix}$

(3) $44 \longrightarrow 4'4$, where 4' denotes a decrease in the amount of fall of the fundamental frequency.

(4) $3 \begin{pmatrix}1\\2\\4\end{pmatrix} \longrightarrow 3' \begin{pmatrix}1\\2\\4\end{pmatrix}$, where 3' denotes a decrease in the amount of rise of the fundamental frequency after the minimum.

In observing the patterns produced when various words of a sequence were stressed, it became apparent that the sandhi rule which changes a Tone 3 into a Tone 2 was quite independent of stress, i.e. the combination 33 became 23 no matter where the stress was located, and when the first Tone 3 was stressed it appeared identical to a stressed Tone 2. The change of Tone 2 into a Tone 1 was not of this type. A Tone 2 became a Tone 1 in the correct environment only when not stressed. When stressed, the Tone 2, which had supposedly changed into a Tone 1, appeared identical to a stressed Tone 2. This is one of the reasons mentioned in Section I for desiring the stress to be marked before the phonological rules function. The sandhi rule converting a Tone 2 into a Tone 1 is therefore sensitive to stress, whereas the rule converting a Tone 3 into a Tone 2

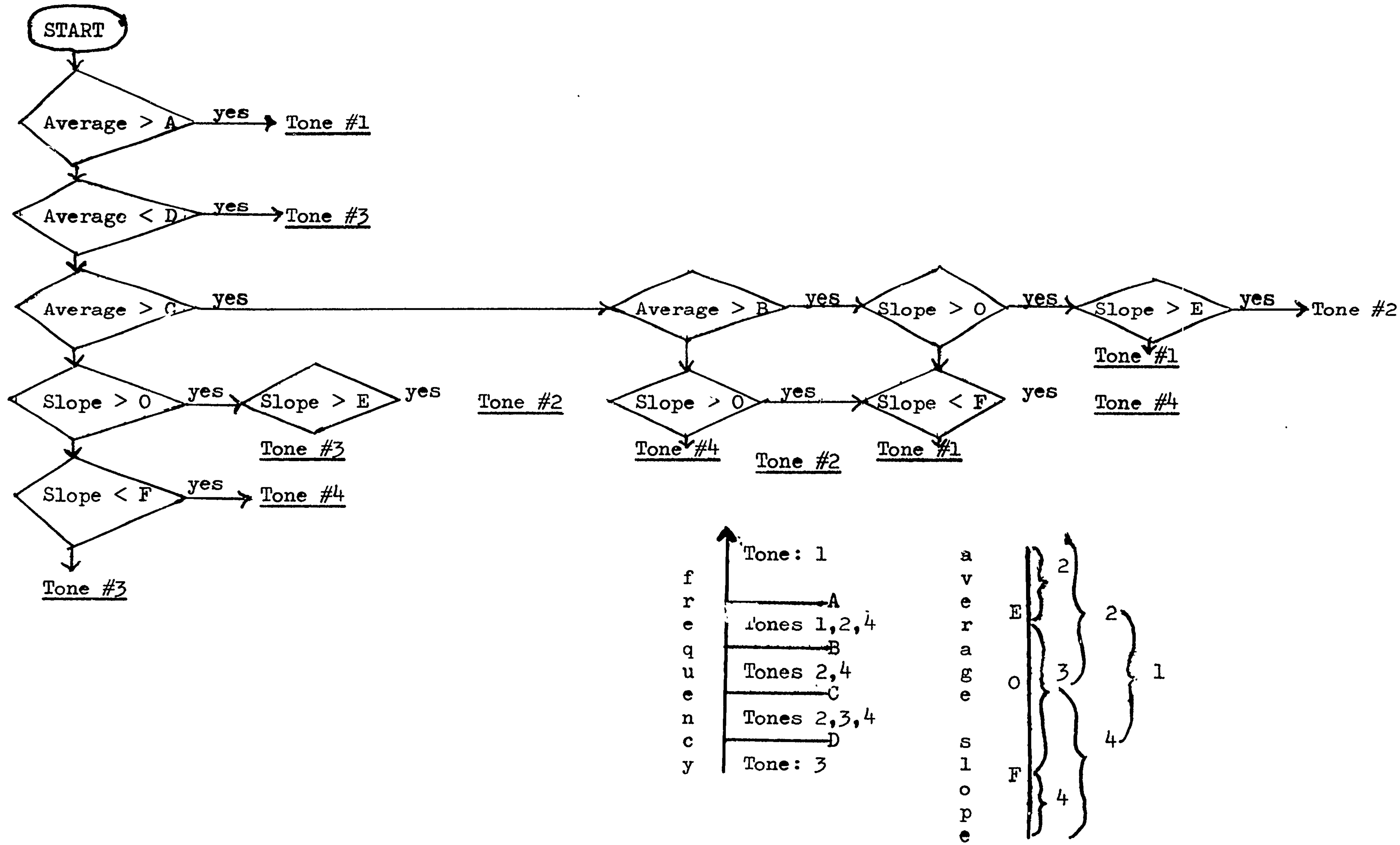


Figure 4. Separation of Tones, Words Spoken in Isolation

Allotones	High / Low	Dynamic / Static	TH1	TH2	TH3	Modifications indicated by the codings
1	+	-	-	-	-	citation form
2	+	+	-	-	-	citation form
3	-	-	-	-	-	citation form
4	-	+	-	-	-	citation form
1'	+	-	+	-	-	average frequency lowered ≈ 5%
2'	+	+	+	+	-	average frequency lowered ≈ 5% average slope raised ≈ 50%
3'	-	-	+	-	-	average frequency raised ≈ 15% average slope lowered ≈ 150%
4'	-	+	+	-	-	average slope lowered ≈ 30%
4''	-	+	+	+	-	average slope lowered ≈ 15%
4'''	-	+	-	-	+	average slope raised ≈ 30%
4^{iv}	-	+	+	-	+	average frequency raised ≈ 5% average slope raised ≈ 30%
4^{v}	-	+	-	+	+	average slope raised ≈ 50%

TH1, TH2, and TH3 are abbreviations of Threshhold 1, etc.

Figure 5. Some Allotones in Mandarin

is not. For the present we will assume that the allotonic rules are not sensitive to stress. When stress has been investigated systematically, we will be able to determine whether this assumption is useful, or whether we should specify all or some allotonic changes as sensitive to stress.

Listed below are some sandhi and allotonic rules which have been derived from the speech of the two informants. "," will be used to denote a pause; "X" will be used to dentoe "no pause." The rules are ordered and repeatable, and operate from left to right starting at a pause. In each rule the number refers to the tone and all of its allotones. All references to changes are to changes with respect to the <u>tone</u>, not with respect to the allotonic state of the tone. Thus, rule #9 does <u>not</u> raise the average slope another 30% above the level obtained by rule #8 functioning.

(1) 3 ----→ 2 in environment (hereafter abbreviated env.) ___3

(2) 2 ----→ 1 in env. $\begin{matrix}1\\2\end{matrix}$ ___X

(3) 1 ----→ 1' in env. $\begin{matrix}3\\4\end{matrix}$ ___. where 1' has its average frequency lowered $\approx$ 5%, and 2' has its average frequency lowered $\approx$ 5%, and its average slope raised $\approx$50%.

(4) 1 ----→ 1' in env. ___4

(5) 3 ----→ 3' in env. ___X where 3' has its average frequency raised $\approx$ 15%, and its average slope lowered $\approx$150%.

(6) 4 ----→ 4' in env. X___. where 4' has its average slope lowered $\approx$30%.

(7) 4 ----→ 4'' in env. 3___. where 4'' has its average slope lowered $\approx$15%.

(8) 4 ----→ 4''' in env. ___X. where 4''' has its average slope raised $\approx$30%.

(9) 4 ----→ 4^{iv} in env. ___1. where 4^{iv} has its average frequency raised $\approx$ 5%, and its average slope raised $\approx$30%.

(10) 4 ----→ 4^{v} in env. ___4. where 4^{v} has its average slope raised $\approx$50%.

The amounts of change in the parameters as indicated by the allotonic rules are of necessity rather vague, since there is considerable variation permissible. The above values merely represent average effects observed from the speech of the two informants.

A rather crude attempt has been made to code the tones and the rules into a binary type coding slightly similar to that proposed by Helen Wong.[11] The main difference between this coding and that proposed by Wong, or that of distinctive features applied to segmentals, is that the plus-minus indicators are not necessarily diametrically opposed. Specifically, the features TH1, TH2, etc. (the TH stands for threshold) are not in diametric opposition but are merely a change of coding. Some simplification of the rules is made possible by using this type of rules, which might be construed to indicate that this approach will be fruitful in the future.

We begin by defining four tones, 1, 2, 3, and 4, which are specified by giving two features, high-low and dynamic-static. From these four tones we generate the 12 allotones indicated in the above rules, by adding change features in a binary coding. Figure 5 gives the entries which will be used in the following manipulations to specify the tones and their allotones as shown.

It has been demonstrated by Halle[12] and Bever[13] that rules may be simplified by introducing variables in the notation. In the rules shown below the variable "a" will be used to indicate whether the place which it occupies should be the same sign or different from that of a previously used "a" in the same rule, e.g. [a high] ----> [-a high] indicates that different signs should occupy the two positions, whereas [a high] ----> [a dynamic] indicates that the same sign (+ or -) should occupy both positions. Once again, "." will be used to denote a pause, and "X" will be used to indicate "no pause." The rules are ordered and repeatable, that is, rule number n should be consecutively applied as long as its conditions are met, before going on to rule number (n + 1). The rules apply from left to right on a morpheme string, starting at a pause.

$$(1)\ \begin{bmatrix} -\text{ high} \\ -\text{ dynamic} \end{bmatrix} \dashrightarrow \begin{bmatrix} +\text{ high} \\ +\text{ dynamic} \end{bmatrix} \text{ in env. } \underline{\qquad\qquad} \begin{bmatrix} -\text{ high} \\ -\text{ dynamic} \end{bmatrix}$$

$$(2)\ [+\text{ dynamic}] \dashrightarrow [-\text{ dynamic}] \text{ in env. } [+\text{ high}]\ \underline{[+\text{ high}]}\ X$$

$$(3)\ [\quad] \dashrightarrow \begin{bmatrix} +\text{ TH1} \\ a\text{ TH2} \end{bmatrix} \text{ in env. } [-\text{ high}]\ \underline{\begin{bmatrix} +\text{ high} \\ a\text{ dynamic} \end{bmatrix}}.$$

$$(4)\ [\quad] \dashrightarrow \begin{bmatrix} -a\text{ TH1} \\ a\text{ TH3} \end{bmatrix} \text{ in env. } \underline{\begin{bmatrix} -\text{ high} \\ a\text{ dynamic} \end{bmatrix}}\ X$$

$$(5)\ [\quad] \dashrightarrow \begin{bmatrix} a\text{ TH1} \\ -a\text{ TH2} \end{bmatrix} \text{ in env. } \underline{\begin{bmatrix} -\text{ high} \\ +\text{ dynamic} \end{bmatrix}} \begin{bmatrix} a\text{ high} \\ -a\text{ dynamic} \end{bmatrix}.$$

$$(6)\ [\quad] \dashrightarrow [+\text{ TH1}] \text{ in env. } \underline{\begin{bmatrix} +\text{ high} \\ -\text{ dynamic} \end{bmatrix}} \begin{bmatrix} -\text{ high} \\ +\text{ dynamic} \end{bmatrix}.$$

$$(7)\ [\quad] \dashrightarrow [+\text{ TH1}] \text{ in env. } X \underline{\begin{bmatrix} -\text{ high} \\ +\text{ dynamic} \end{bmatrix}}.$$

$$(8)\ [\quad] \dashrightarrow [+\text{ TH2}] \text{ in env. } \begin{bmatrix} -\text{ high} \\ -\text{ dynamic} \end{bmatrix} \underline{\begin{bmatrix} -\text{ high} \\ +\text{ dynamic} \end{bmatrix}}.$$

Rule (1) is a sandhi rule which changes Tone 3 into Tone 2 when it precedes another Tone 3. Rule (2) is a sandhi rule which changes Tone 2 into a Tone 1 when it follows either Tone 1 or Tone 2 and does not precede a pause. Rule (3) is an allotonic rule which modifies the average frequency of Tones 1 and 2 and raises the slope of Tone 2 when either tone follows a Tone 3 or a Tone 4. Rule (4) is an allotonic rule which modifies the average slope of Tone 4 and changes the average frequency and the average slope of Tone 3 when either tone does not precede a pause. Rule (5) is an allotonic rule which modifies the average frequency and average slope of Tone 4 by amounts which depend on whether the Tone 4 precedes a Tone 1 or another Tone 4. Rule (6) is an allotonic rule which modifies Tone 1 in the same manner when it precedes a Tone 4 as it was modified when following either Tone 3 or Tone 4. Rule (7) is an allotonic rule which changes the average slope of a Tone 4 when it precedes a pause. Rule (8) is an allotonic rule which further changes the slope

of a Tone 4 if it follows Tone 3 when preceding a pause. A correlation between the two sets of rules, in terms of which rule is replaced by which rule, and how many binary bits are saved in the process, is given below.

The first rule number mentioned refers to the binary rules, and the second number refers to the previous set of rules.

Rule (1) replaces Rule (1), there is no saving of information.

Rule (2) replaces Rule (2), there is a saving of 4 bits in the binary rule.

Rule (3) replaces Rule (3), there is a saving of 9 bits in the binary rule.

Rule (4) replaces Rules (5) and (8), there is a saving of 8 bits in the binary rule.

Rule (5) replaces Rules (9) and (10), there is a saving of 12 bits in the binary rule.

Rule (6) replaces Rule (4), there is a saving of 2 bits in the binary rule.

Rule (7) replaces Rule (6), there is a saving of 4 bits in the binary rule.

Rule (8) replaces Rule (7), there is a saving of 4 bits in the binary rule.

Thus, there is a saving of 44 bits when the binary rules are used in place of the previous rules.

As is immediately apparent from Figure 5, the coding of the allotones is quite arbitrary with relation to the binary bits, TH1, TH2, and TH3. This arbitrariness indicates that the subtle relations between tones have not yet been discovered. Until such time as one can specify contrastive binary bits which have physical meaning in their own right, and also facilitate the writing of the rules, it must be assumed that one is working with rather artificial codings, which will not reveal the common features of tone languages, or show how these features interact in producing the effects of sandhi and allotonic change due to environment.

What is lacking, then, is a distinctive features type framework with which to describe the fundamental frequency parameter of a tone language.

The sandhi change of a Tone 3 into a Tone 2 is evidenced by a change in the average slope of the Tone 3 in the correct environment to that of a similarly situated Tone 2. The average slope after this change is quite different from the average slope of a similarly situated natural Tone 2 for one of the informants, but quite similar for the other. If the differences were relatively constant, then there would be a chance to specify the original tones in a recognition procedure with one less ambiguity.

The other sandhi change, that of a Tone 2 into a Tone 1 in certain environments, is apparently quite complete for one of the informants. The other informant, however, has the average frequency of a derived Tone 1 quite far below that of a natural Tone 1. This informant also retains a good deal of the upward slope of the Tone 2 when it becomes a Tone 1, so that one might seriously question whether this change is actually to be included under the classification of shandh, or whether it should be referred to as merely an allotonic change in which the allotone closely resembles a Tone 1 in the same environment. The former course was chosen because it is the one specified by Chao; the evidence was not strong enough in this brief sample to challenge his decision.

The neutral tone has been omitted from consideration at present, although a small amount of data on it has been collected. Upon examination of this data, it appeared that the introduction of the neutral tone also introduced some stress in the surrounding tones, as evidenced by a marked rise in the average frequencies of preceding and following tones. While this effect may have been due purely to an unwise choice of phrases which contained the neutral tone, or to an inadvertent change in the speech of the informants due to reading phrases which were not nonsensical, it seems reasonable that we should expect stress to occur when "non-stress" is introduced.[14] It therefore appears likely that the study of the neutral tone will have to be concurrent with the study of stress. An additional problem in dealing with a neutral tone will be that of setting the criteria for "neutrality," i.e. when is a tone no longer in possession of characteristics which might differentiate it from other tones?

Pike [15] has discussed this problem of tone neutralization and has set a criterion for determining when a tone is neutral. This criterion might be simply stated by saying that a tone becomes neutral when one cannot distinguish it from all of the other tones in the language. Or, to state it another way, a given tone is a neutral tone when it is impossible to identify exactly which of the basic tones of the language it is supposed to be. The difficulty in applying this criterion is that it assumes knowledge of exactly what variations are permissible for a given tone in a given environment, i.e. the tone is not neutral just because one cannot distinguish which tone it is, unless the recognition system used is perfect. This approach, however, is not to be ignored simply because one needs to assume a perfect analysis system. One can just as well use an empirical approach, by setting up the areas of neutrality to the best of one's knowledge, and then working to shrink these areas of ambiguity by improving the recognition scheme.

Another factor worthy of comment is that such a criterion for neutrality will introduce a whole host of neutral tones in place of the one neutral tone which is assumed in the classical analyses of Mandarin. That is, one can have neutral tones which are ambiguous as to which of two tones the tone in question is supposed to be, or the ambiguity can be among three tones, etc. If we define the order of neutrality to be #1 when the ambiguity is between two tones, and to be #2 when the ambiguity is among three tones, and so forth, then we have a set of neutral tones in each of the orders possible. For a tone language with n contrastive tones we can write an expression for the maximum number of neutral tones which can occur in each order as follows:

$$(\text{ORDER \#1})_{max} = C^n_2 = \frac{n\cdot(n-1)}{2!}$$

$$(\text{ORDER \#2})_{max} = C^n_3 = \frac{n\cdot(n-1)\cdot(n-2)}{3!}$$

$$(\text{ORDER \#3})_{max} = C^n_{B+1} = \frac{n\cdot(n-1)\cdot(n-2)\ .\ .\ .\ (n-B)}{(B+1)!}$$

Thus, the maximum number of neutral tones in the language is given by:

$$(\#\text{ of neutral tones})_{max} = C_2^n + C_3^n + \dots + C_n^n =$$

$$\frac{n\cdot(n-1)}{2!} + \frac{n\cdot(n-1)\cdot(n-2)}{3!} + \dots + \frac{n\cdot(n-1)\cdot(n-2)\dots(n-n+1)}{n!}$$

While it is improbable that there will be the maximum number of contrastive neutral tones of a given order except for the order #(n-1), this type of approach is still quite different from the classical approach, in which all tones which become neutralized become completely neutralized, i.e. of the order #(n-1).

From the standpoint of a generative system, these various orders of neutrality need not concern us, since we merely specify the possible variations for each tone as a function of its environment, and do not worry about whether or not another tone's permissible variations intersect with these specifications. (In the classical treatment of Mandarin, however, the generation of the neutral tone is not determined by tone and environment, but rather by lexical information and environment.) From the recognition standpoint, however, we must be conscious of the fact that we can have several levels of ambiguity, and several ambiguities at each level.

Appendix 4 gives the values of some of the extracted parameters as a function of environment, for the reader's information.

V. Summary

The form of a generative system for fundamental voice frequency in Mandarin is discussed, and several assumptions are made in order to reduce the problem to a manageable size. Data was gathered from two speakers, by having them read a prepared list which contained isolated words, two-tuples, and three-tuples, in all possible combinations of the four basic tones. The data gathering system utilized a modified "Vocoder" pitch extractor, a two-channel graphical recorder, and an electronic digital computer, which was used to plot out the pitch curve, and also to extract various parameters of the pitch contour. Examples were given of the four basic tones of Mandarin, and, on the basis of the data gathered, a set of rules was proposed which would account for the tone sandhi and certain allotonic changes evident in Mandarin speech.

Footnotes

[1]W. S-Y. Wang, "Mandarin Phonology," POLA No. 6, The Ohio State University Research Foundation, 1963, p. 4.

[2]"Intrinsic pitch" refers to the phenomenon of pitch changing as a function of the vowel being articulated, i.e. higher vowels have higher pitch. This effect carries no information, since it is predictable.

[3]See Yao, Shen, et al. "Some Spectrographic Light on Mandarin Tone 2 and Tone 3," pp. 265-70 in Study of Sounds IX, Phonetic Society of Japan, 1961, for a synopsis of the ideas concerning Mandarin tones in isolation.

[4]Chao, Yuen Ren, "Tone and Intonation in Chinese," a paper read before the 145th meeting of the American Oriental Society, April 20, 1933. Also appeared in Bulletin of the Institute of History and Philology, 4 (1933), pp. 121-34.

[5]See Appendix 1 for the linguistic backgrounds of the two informants and a list of the test materials read.

[6]See Appendix 2 for block diagram and response of the pitch extractor and the amplitude tracker.

[7]See Appendix 3 for the formulas used to define the quantities listed.

[8]See Hockett, C.F., "Peiping Morphophonemics," Language XXVI (1950), pp. 63-85, and Martin, S.E., "Problems of Hierarchy and Indeterminacy in Mandarin Phonology," Bulletin of the Institute of History and Philology, 29, Part I, pp. 209-30.

[9]Li, K-P., "Tone Perception Experiment," POLA No. 6, The Ohio State University Research Foundation, 1963, pp. 19-26.

[10]Chao, Y.R., Mandarin Primer, Cambridge, Mass.: Harvard University Press, 1948, pp. 107-13.

[11]Wong, Helen, "Outline of the Mandarin Phonemic System," Word 9.3 (December, 1953), pp. 268-76.

[12]Halle, M., "A Descriptive Convention for Treating Assimilation and Dissimilation," Quarterly Progress Report No. 66 (July 15, 1962), Research Laboratory of Electronics of Massachusetts Institute of Technology, pp. 295-6.

[13]Bever, T.G., "Formal Justification of Variables in Phonemic Cross-Classifying Systems," Q.P.R. No. 69 (April 15, 1963), R.L.E. of M.I.T., pp. 200-202.

[14] Chao, Y.R., "Tone and Intonation in Chinese," a paper read before the 145th meeting of the American Oriental Society, April 20, 1933. Also appeared in Bulletin of the Institute of History and Philology, 4 (1933), pp. 121-34.

[15] Pike, Kenneth L., "Operational Phonemics in Reference to Linguistic Relativity," Journal of Acoustical Society of America, 24 (1952), p. 622.

Appendix 1

A. Monosyllables

疤 叉 媽 發
拔 茶 麻 罰
把 蹅 馬 法
霸 詫 罵 琺

B. Two-tuples

疤 叉 媽 發
拔 茶 麻 罰
把 蹅 馬 法
霸 詫 罵 琺

C. Three-tuples

疤 叉 媽
拔 茶 麻
把 蹅 馬
霸 詫 罵

D. Neutral Tones

三個
一個
七個
八個
十個
五個
九個
四個
二個
六個
一是一，二是二
說得來
聽不懂
中國話
難得多
誰的筆

學不會
好的多
兩個人
想得很
嗓子大
到了家
就是難
豆腐腦
看得見
飛來了
王先生
椅子呢（好的了）
对了吧
高不高
買不買
貴不貴
行不行

E. Stressed Phrases

我的紅鉛筆

名望很大

汽車賣了

有錢的人

這個收音機

喝酒的人

聰明小子

紅紅的臉

櫻桃小口

黃種人

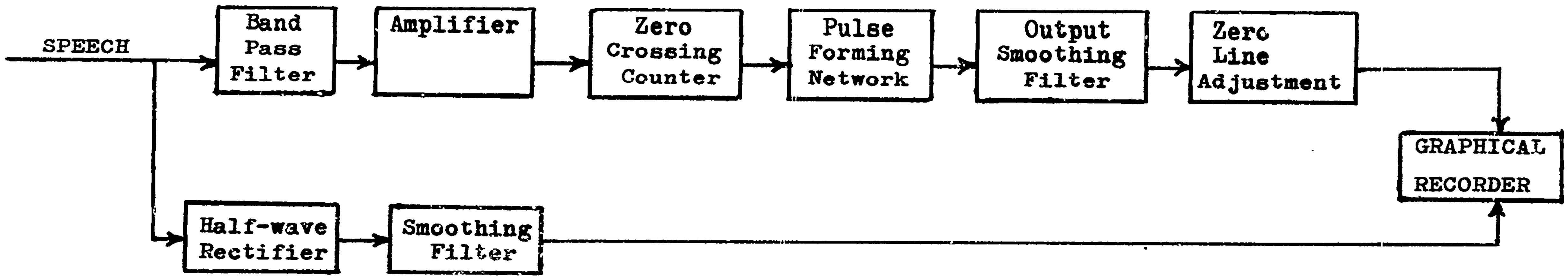
冷氣壞了

Appendix 2

Linguistic Backgrounds of the Two Informants
and
List of Test Materials Read

Informant	# 1	# 2
Name	Kung-Pu Li	Lillian Liu
Sex	Male	Female
Age	29	26
Dialect	Mandarin	Mandarin
Place of Birth	Shanghai	Shanghai
Elementary School	Peking	Shanghai, Nanking
High School	Peking, Taiwan	United States
College	Taiwan, United States	United States
Dialect spoken by parents	Fukianese, Mandarin	Shanghai, Kiangsi, Mandarin
Dialect spoken at home	Mandarin	Shanghai, Kiangsi, Mandarin
Other dialects	---	Cantonese
Proficiency in other languages	English, fair	English, fluent Spanish, fluent

PITCH EXTRACTOR



AMPLITUDE TRACKER

Overall system response: Rise Time of Pitch Extractor = 15 msec
Rise Time of Amplitude Tracker = 30 msec

Appendix 3

Block Diagram Analysis Scheme

Appendix 4

Definition of Parameters Extracted from Pitch Curves

R = Number of data points - 1.

$$\text{Average Frequency} = \frac{\sum_{n=0}^{R} \text{Frequency}_n}{R+1}$$

Frequency of Minimum = Frequency of that point whose frequency is lower than the point preceding it and lower than or equal to the frequency of the following point.

$$\text{Slope toward minimum} = \frac{\text{Minimum frequency-Initial frequency}}{\text{Number of 40-msec intervals between initial and minimum.}}$$

$$\text{Slope from minimum} = \frac{\text{Final frequency-Minimum frequency}}{\text{Number of 40-msec intervals between minimum and final point.}}$$

$$\text{Starting slope} = \frac{\text{Frequency of 3rd data point-Frequency of initial point}}{2}$$

$$\text{Average slope} = \frac{\text{Frequency of final point-Frequency of initial point}}{\text{Number of 40-msec intervals between initial and final points.}}$$

Relative slope = Average slope-Initial slope.

Concavity was called upward if the relative slope was positive, and called downward if the relative slope was negative.

The magnitude of the relative slope serves as a measure of concavity.

Appendix 5

Acoustical Data Extracted

A. Data for Isolated Tones

Informant # 1	average	maximum	minimum
Tone 1:			
average frequency	136.5	145.4	129.2
average slope	1.52	3.34	0.56
Tone 2:			
average frequency	113.0	117.2	109.0
average slope	4.96	6.68	3.34
no minimums are evident			
Tone 3:			
average frequency	90.8	95.5	87.6
average slope	1.36	3.58	-0.37
average minimum	80.4	85.0	71.7
Tone 4:			
average frequency	117.0	125.1	109.2
average slope	10.6	7.4	13.36

Informant # 2	average	maximum	minimum
Tone 1:			
average frequency	231.9	242.9	220.6
average slope	1.09	4.38	-1.67
Tone 2:			
average frequency	195.0	202.1	190.5
average slope	3.83	6.88	0.42
average minimum	187.1	195.0	185.0
Tone 3:			
average frequency	151.8	159.1	145.8
average slope	-5.65	-3.00	-14.17
average minimum	142.0	152.5	135.0
Tone 4:			
average frequency	198.2	218.5	182.5
average slope	-17.1	-11.7	-23.8

Appendix 5

Acoustical Data Extracted

B. Data for Tones in Sequence

		Informant # 1			Informant # 2		
		average	maximum	minimum	average	maximum	minimum
Tone 1:							
env. 1____.	AV	130	136	124	246	269	234
	AVSL	-.38	.42	-2.09	-.84	1.5	-8.13
env. 2____.	AV	127	131	122	246	263	233
	AVSL	.50	2.23	-1.67	-.45	5.0	-4.07
env. 3____.	AV	123	135	117	233	244	211
	AVSL	1.55	4.45	-1.25	-1.21	7.5	-31.7
env. 4____.	AV	122	147	104	231	256	194
	AVSL	.86	3.34	-2.09	.75	6.25	-2.50
env. ____1.	AV	128	135	122	241	265	226
	AVSL	1.24	5.01	-.84	2.59	6.5	0
env. ____2.	AV	125	133	120	240	257	226
	AVSL	1.11	5.85	-2.0	.92	8.1	-7.5
env. ____3.	AV	125	137	121	244	259	232
	AVSL	1.59	4.17	.33	3.14	8.3	-2.4
env. ____4.	AV	122	129	114	237	253	224
	AVSL	2.38	5.84	1.25	1.72	8.1	-3.2
Tone 2:							
env. 1____.	AV	111	120	105	191	203	181
	AVSL	2.57	5.01	-.42	.64	7.5	-4.38
	MIN	99	101	95	180	190	172
env. 2____.	AV	111	118	104	189	205	176
	AVSL	2.24	4.45	-.84	2.16	7.5	-2.9
	MIN	103	112	98	178	190	165
env. 3____.	AV	105	111	99	177	190	165
	AVSL	3.96	7.24	-1.34	7.43	15.0	3.7
	MIN	95	102	88	166	178	160
env. 4____.	AV	104	111	92	182	190	170
	AVSL	2.60	5.43	0	6.74	9.0	1.7
	MIN	97	105	90	173	180	167

		Informant # 1			Informant # 2		
		average	maximum	minimum	average	maximum	minimum
env. $\frac{1}{2}$ ____ 1.	AV	118	122	114	216	228	207
	AVSL	.94	8.35	-2.92	1.58	5.0	-5.4
	MIN	116	118	111	207	210	205
env. $\frac{1}{2}$ ____ 2.	AV	123	125	121	214	223	205
	AVSL	.21	3.34	-2.92	4.03	10.0	-1.7
	MIN	121	123	120	206	210	203
env. $\frac{1}{2}$ ____ 3.	AV	117	126	90	222	237	212
	AVSL	3.04	12.11	-1.34	6.05	10.8	-0.4
	MIN	104	125	68	212	222	205
env. $\frac{1}{2}$ ____ 4.	AV	120	128	120	207	231	191
	AVSL	.58	2.34	0	2.75	6.7	-2.5
	MIN	105	105	105	203	215	192
env. $\frac{3}{4}$ ____ 1.	AV	111	115	105	206	223	191
	AVSL	4.26	7.51	1.67	11.7	14.5	6.5
	MIN	108	112	105	195	195	195
env. $\frac{3}{4}$ ____ 2.	AV	115	118	110	199	218	184
	AVSL	4.57	7.51	2.92	9.32	14.6	3.8
	MIN	---	---	---	207	207	207
env. $\frac{3}{4}$ ____ 3.	AV	113	117	110	214	235	195
	AVSL	4.52	7.24	1.95	13.2	20.3	5.4
	MIN	110	110	110	---	---	---
env. $\frac{3}{4}$ ____ 4.	AV	112	119	106	190	201	181
	AVSL	5.56	8.35	1.25	9.36	13.3	4.1
	MIN	110	110	110	188	188	188
Tone 3:							
env. 1 ____.	AV	85	98	77	162	174	152
	AVSL	-1.53	1.95	-6.12	-6.72	0	-12.8
	MIN	75	87	68	150	164	142
env. 2 ____.	AV	85	104	76	162	176	151
	AVSL	-1.21	1.0	-4.34	-6.73	.4	-12.1
	MIN	79	102	70	149	155	142
env. 3 ____.	AV	88	94	78	159	183	132
	AVSL	-1.97	-0.28	-5.01	-5.98	0	-12.9
	MIN	78	87	70	148	159	125
env. 4 ____.	AV	82	87	72	158	165	146
	AVSL	-0.66	2.39	-3.34	-1.41	5.8	-6.7
	MIN	72	78	68	150	155	142

		Informant # 1			Informant # 2		
		average	maximum	minimum	average	maximum	minimum
env. ___1.	AV	90	105	75	179	214	160
	AVSL	-4.83	.28	-9.18	-6.62	-1.9	-12.2
	MIN	80	93	73	158	170	142
env. ___2.	AV	92	102	85	179	199	161
	AVSL	-5.97	-0.67	-13.36	-9.7	-2.5	-17.1
	MIN	83	90	78	163	170	155
env. ___3.	AV	118	124	112	216	230	188
	AVSL	2.54	8.35	-1.39	5.1	11.7	-8.8
	MIN	116	123	107	212	215	210
env. ___4.	AV	90	100	78	170	195	153
	AVSL	-4.42	10.02	-15.03	-5.38	-0.7	-9.8
	MIN	83	92	72	157	165	147

Tone 4:

env. 1___.	AV	115	122	102	217	233	195
	AVSL	-11.6	-5.34	-19.48	-27.25	-19.2	-37.4
env. 2___.	AV	114	121	109	215	239	183
	AVSL	-13.35	-6.68	-17.81	-25.2	-13.5	-31.7
env. 3___.	AV	112	122	105	215	232	206
	AVSL	-10.2	-2.78	-18.37	-23.4	-14.6	-32.5
env. 4___.	AV	114	125	97	219	234	207
	AVSL	-11.9	-4.68	-17.53	-20.6	-9.4	-35.4
env. ___1.	AV	119	125	114	234	255	215
	AVSL	-6.15	-1.34	-11.69	-9.69	-4.2	-17.5
env. ___2.	AV	115	121	110	221	237	203
	AVSL	-5.94	0	-8.35	-13.8	-5.6	-26.9
env. ___3.	AV	115	125	108	225	243	212
	AVSL	-6.93	-3.67	-10.58	-11.78	.36	-22.5
env. ___4.	AV	118	126	110	223	238	207
	AVSL	-3.30	1.67	-7.93	-9.88	-2.1	-17.9